The overall goal of this project is to promote the accessibility and dissemination of biomedical information so that the research community can better leverage existing knowledge. We have a particular emphasis on illuminating biomedical "dark data." By analogy to the dark matter that is unaccounted for in the universe, dark data is defined by being unseen or underutilized by the scientific community. This project specifically focuses on making these dark data resources **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (**FAIR**) [ref].

During the first project period of this grant, we introduced BioGPS (http://biogps.org), a gene portal for aggregating information on human genes and proteins. BioGPS illuminates dark data by creating a simple platform to discover and access gene-centric websites. Targeted at non-computational biologists, BioGPS emphasizes user customizability and community extensibility. Currently, BioGPS is used by over 150,000 unique visitors per year generating almost two million page views. During the second project period, we added MyGene.info (http://mygene.info), a new standalone project created by abstracting a key part of BioGPS data management system. Targeted at the bioinformatics community, MyGene.info integrates gene and protein annotation data into a simple and high performance web service endpoint. MyGene.info illuminates dark data on gene and protein annotations by pre-integrating over 200 annotation types. MyGene.info averages 4000 unique users and 10 million accesses per month.

This renewal proposal includes four Specific Aims that build on this track record and long-standing interest in illuminating biomedical dark data. The first two aims address the continued development and evolution of our existing resources -- BioGPS (**Aim 1**) and MyGene.info (**Aim 2**). The second half of the proposal builds on this foundation to introduce new tools for bioinformaticians (**Aim 3**) and the genomic research community (**Aim 4**).

Aim 1: Integrate BioGPS with community tools for plugin, gene list, and data set management. Utilizing specialized services strengthens each individual tool and the overall biomedical Big Data ecosystem.

A) Leverage the BD2K AZTEC resource repository as the BioGPS plugin library. This collaboration will serve as a key use case for AZTEC development and a stronger long-term foundation for BioGPS.
B) Integrate the Tribe repository for BioGPS gene list management. Collaborating with this dedicated service will allow us to incorporate new functionality while increasing focus on core BioGPS features.
C) Import differential expression data from GEO and ArrayExpress. This feature will enable searching for gene expression experiments in which a gene of interest was differentially expressed.

Aim 2: Expand MyGene.info to include additional highly-requested annotation sources. This expansion will include data from a major data repository, as well as from smaller domain-specific data resources.

A) Expand the species and annotation coverage available through EnsemblGenomes. This will include annotations for thousands of species available from Ensembl Bacteria, Ensembl Plants and more.
B) Expand the gene annotations from a dozen of specialized data resources. Importing these data will improve their accessibility and visibility to the community, as well as the utility of MyGene.info.

Aim 3: Generalize the MyGene.info software pattern to other biomedical entity types. As the number of biomedical data resources increases, illuminating dark data becomes relevant to other domain areas.

A) Develop a generic Biothings SDK. This software development kit will allow any developer to easily create a new web service resource based on the high performance MyGene.info platform.
B) Build the drug and chemical equivalent of MyGene.info. Annotations of drug/chemical properties are highly fragmented across resources, creating a need for high-performance integrated web services.
C) Build the disease equivalent of MyGene.info. We will create a web service endpoint that integrates annotations of disease properties (including links to associated genes and phenotypes).

Aim 4: Create BioReel, a new application for monitoring updates about Biothings entities. BioReel users can create customized alerts for new information on specific genes, diseases, drugs, and variants of interest.

A) Build infrastructure to detect and store entity-specific changes within the Biothings SDK. The Biothings import modules will be extended to compare each update with the preceding version.
B) Create interfaces for users to register and receive updates. Users will be able to access chronological updates through a web page, mobile app, email alerts, and RSS feed.